MPG against Transmission

Sean Dunagan - October 23, 2015

Executive Summary

As detailed below, there does appear to be a difference in MPG between Automatic and Manual transmission types. Manual transmission types appear to provide better MPG values when accounting for other relevant variables such as the Weight of the car, the number of Cylinders of the car, and the number of Carburetors. The estimated difference in MPG is about 1.78 more miles per gallon, obtained with 75.1% confidence via a linear model fitting MPG against Transmission Type, Weight, Cylinders and Carburetors.

Exploratory Data Analyses

A natural first step in addressing these questions would be to see the distribution of MPG values for each transmissions type in the dataset we are using. Figure 1 in the Appendix (MPG by Transmission Type) displays this visually via a boxplot. It is easy to see via this boxplot that the MPG values tend to be higher for the Manual transmission type. The quantile data displayed in the boxplot is given below via the R summary() function for Automatic and Manual transmissions respectively

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.40   14.95   17.30   17.15   19.20   24.40
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   15.00   21.00   22.80   24.39   30.40   33.90
```

A natural next step would be to look at all of the variables contained in the mtcars dataset

```
##           mpg cyl disp  hp drat   wt  qsec vs am gear carb
## Mazda RX4  21   6  160 110  3.9 2.62 16.46  0  1    4    4
```

Regression Models

An obvious model to explore would be the interaction between the transmission type and the MPG values. The coefficients for this linear model are given in Figure 7 of the Appendix (According to the ?mtcars R documentation, the values for am are 0 = automatic, 1 = manual):

We can see from the coefficients in the figure that the estimate for the intercept is 17.147368; this represents the estimated MPG value for Automatic Transmissions. The estimate for the am coefficient is 7.244939. This indicates that for Manual Transmissions the estimated MPG value is 17.147368 + 7.244939; this yields an estimated MPG value of 24.39231 for Manual Transmissions. This also indicates that the difference in MPG values between transmission types is 7.244939 MPG.

While this displays the relationship between MPG and transmission type, anyone with subject matter knowledge in the automotive realm will note that there are a number of additional variables which are significantly related to the MPG of a car (and which are not redundant to transmission type). To this note, the R^2 value of the mpg ~ am model is only 0.3598, meaning that only 35.98% of the variance in the data is explained by our mpg ~ am model. Surely there are other variables affecting the response (MPG).

In particular, the number of cylinders and the weight of a car are very related to the MPG of a car. As such, we should inspect a model which includes these variables; Figure 4 in the Appendix displays the statistical data for this model.

The R^2 value for this model is 0.8303, MUCH higher than the 0.3598 value for the mpg ~ am model; this indicates that our new model accounts for 47% more of the total variance in MPG in the mtcars dataset than the previous model did. In terms of the regressors' coefficients, the model indicates that both cyl and wt have an inverse relationship with MPG, meaning that an increase in the amount of cylinders or an increase in the weight of the car will decrease the MPG of the car. The P values for these two variables are each under .002, so we can conclude this with over 99%

certainty. However, for am, the P value is .893; looking at the standard error for am, we see that it is 1.3, greater than the estimate of 0.1765 for the coefficient. As such, we can't say with much confidence at all whether mpg is inversely or directly relatred to am. As such, we might look for another variable to include in the model to reduce the standard error for am, since this is the variable we are concerned with in particular.

Carburetors are responsible for fuel injection in the car's engine; as such, it would be very reasonable to assume that the number of carburetors is related to the MPG of a car. Figure 5 in the Appendix displays the statistics for such a model. When we add this regressor to the model, we see the P value for the am regressor plummet to .249; at this point, we would say with 75.1% confidence that MPG increases as am increases (meaning with manual transmission). The estimate of 1.78 would suggest that Manual transmissions result in an average of 1.78 more miles per gallon for a car than automatic transmissions. Naturally, as we include more regressors into the model, the actual standard error of the other regressors increases. This is evidenced by the fact that the P values for the wt and cyl coefficients have increased by 609% and 733% respectively. However, given that these P values are still under .02, they are still within the typical 95% confidence interval.

Figure 6 in the Appendix shows a comparison of the 3 models discussed above. As we can see from the data, the Residual Sum of Squares decreases greatly with the addition of the wt, cyl, and carb regressor variables.

Answering the Question of Interest

Given our model taking into account mpg ~ am + wt + cyl + carb, we obtained a regression coefficient of 1.78 for am, with standard error 1.51 and p value .249. As such, the coefficient is above 0 within the bounds of the standard error. Additionally, with a p value of .249, we would say with 75.1% confidence that Manual Transmissions lead to higher MPG values than Automatic transmissions. We would estimate that Manual Transmissions result in an average of 1.78 MPG greater than Automatic Transmissions.

Residual Plot

Figure 2 in the Appendix shows a residual plot. The X-axis on this plot is the residuals of model am ~ wt + cyl + carb; the Y-axis of this plot is the residuals of model mpg ~ wt + cyl + carb. This plot shows us the relationship between MPG and Transmission Type having factored out the other regressor variables wt, cyl and carb. We can see the fit of this plot showing an upward slope, meaning MPG increases from Automatic to Manual transmission type. In figure 8 in the Appendix, we can see the estimate for this slope by constructing a linear fit model for the residual plot and looking at the coefficients for the eAmAgainstWeightCylinderCarbs estimate. This is the same estimate as the am coefficient in the mpgByAmAndWeightAndCylindersAndCarbsModel shown above, which is what we would expect.

Diagnostics

We can run the dffits diagnostic to find some points which are outliers in the data. Figure 3 shows a plot of Residuals vs Fitted. We can see the three cars labeled in the plot as having large absolute diffts values below; this indicates that the predicted response in our model would be changed greatly if one of the cars listed below was removed. These cars are considered outliers in terms of our model.

```
## Chrysler Imperial    Toyota Corolla          Fiat 128        Datsun 710
##         1.0598464         0.9383282         0.8706036        -0.8758761
```

Quantifying the Uncertainy and performing Inference

This was done in sections "Answering the Question of Interest" and "Regression Models" above

Appendix

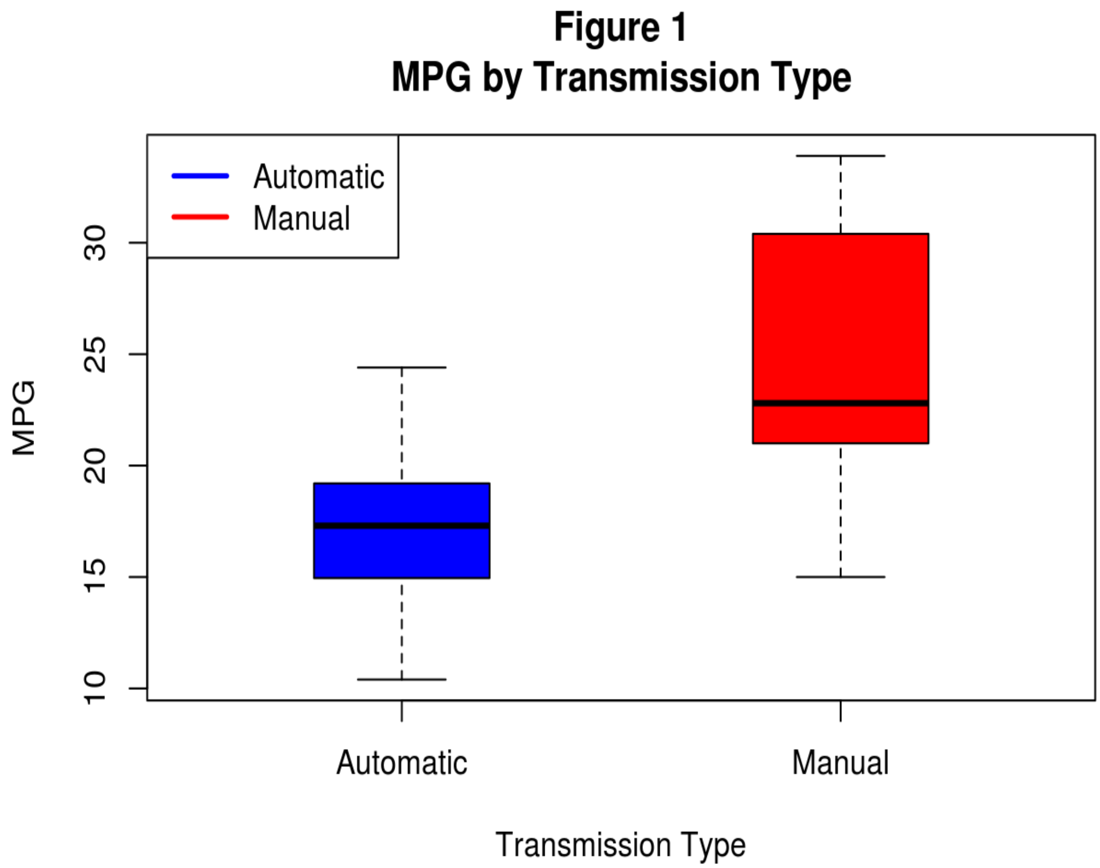

Figure 1
MPG by Transmission Type
Automatic
Manual
MPG
10
15
20
25
30
Automatic
Manual
Transmission Type

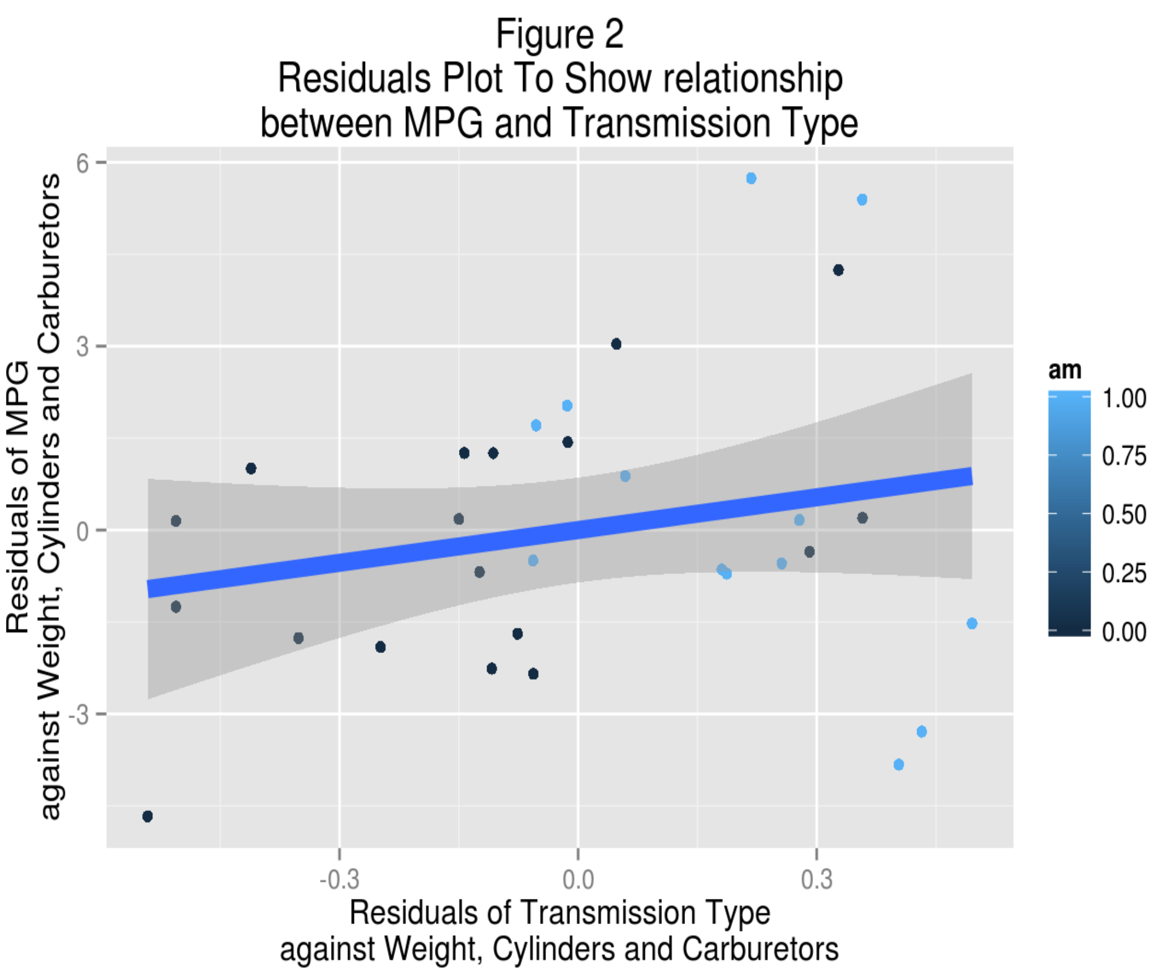

Figure 2
Residuals Plot To Show relationship
between MPG and Transmission Type
Residuals of MPG
against Weight, Cylinders and Carburetors
6
3
0
-3
-0.3
0.0
0.3
Residuals of Transmission Type
against Weight, Cylinders and Carburetors
am
1.00
0.75
0.50
0.25
0.00

Figure 3

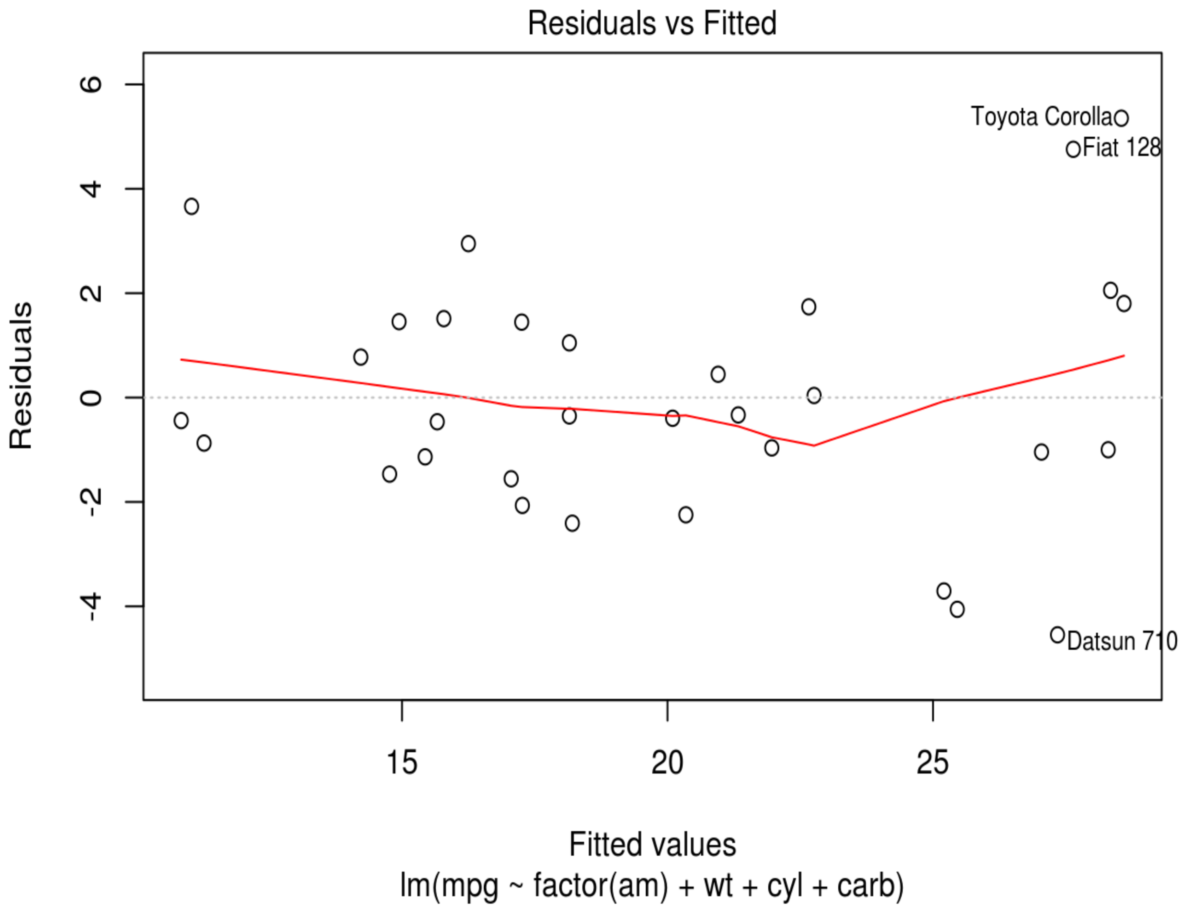

Figure 4

```
summary(mpgByAmAndWeightAndCylindersModel)$coefficients;
```

```
##                Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) 39.4179334  2.6414573 14.9227979 7.424998e-15
## factor(am)1  0.1764932  1.3044515  0.1353007 8.933421e-01
## wt          -3.1251422  0.9108827 -3.4308942 1.885894e-03
## cyl         -1.5102457  0.4222792 -3.5764148 1.291605e-03
```

```
summary(mpgByAmAndWeightAndCylindersModel)$r.squared;
```

```
## [1] 0.8303383
```

Figure 5

```
summary(mpgByAmAndWeightAndCylindersAndCarbsModel);
```

```
##
## Call:
## lm(formula = mpg ~ factor(am) + wt + cyl + carb, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5451 -1.2184 -0.3739  1.4699  5.3528
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  36.8503     2.8694  12.843 5.17e-13 ***
## factor(am)1   1.7801     1.5091   1.180   0.2485
## wt           -2.4785     0.9364  -2.647   0.0134 *
## cyl          -1.1968     0.4368  -2.740   0.0108 *
## carb         -0.7480     0.3956  -1.891   0.0694 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.5 on 27 degrees of freedom
## Multiple R-squared:  0.8502, Adjusted R-squared:  0.828
## F-statistic:  38.3 on 4 and 27 DF,  p-value: 9.255e-11
```

Figure 6

```
anova(mpgByAmModel, mpgByAmAndWeightAndCylindersModel, mpgByAmAndWeightAndCylindersAndCarbsModel);
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(am)
## Model 2: mpg ~ factor(am) + wt + cyl
## Model 3: mpg ~ factor(am) + wt + cyl + carb
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     28 191.05  2    529.85 42.3991 4.672e-09 ***
## 3     27 168.71  1     22.34  3.5756   0.06941 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 7

```
summary(mpgByAmModel)$coefficients;
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## factor(am)1  7.244939   1.764422  4.106127 2.850207e-04
```

Figure 8

```
summary(resdiaulsModel)$coefficients;
```

```
##                                    Estimate Std. Error       t value
## (Intercept)                   -1.177569e-16  0.4192076 -2.809036e-16
## eAmAgainstWeightCylinderCarbs  1.780135e+00  1.4316920  1.243378e+00
##                                Pr(>|t|)
## (Intercept)                   1.0000000
## eAmAgainstWeightCylinderCarbs 0.2233579
```